Advanced Signal Processing Assignment

Yi Zhou 6567008

ASP Assignment 6567008 Yi Zhou

$$\begin{matrix} a & b & c & d & f \\ 3 & 2 & 4 & 3 & 2 \end{matrix} \Rightarrow \mu_1 = \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}, \quad \mu_2 = \begin{bmatrix} a+d \\ b+d \end{bmatrix} = \begin{bmatrix} 3+3 \\ 2+3 \end{bmatrix} = \begin{bmatrix} 6 \\ 5 \end{bmatrix}, \quad \Sigma_1 = \Sigma_2 = \Sigma = \begin{bmatrix} c & f \\ f & c \end{bmatrix} = \begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix}$$

Q1. Take the eigen decomposition of the covariance matrix

$|\lambda I - \Sigma| = 0 \quad \lambda I - \Sigma = \begin{bmatrix} \lambda-4 & -2 \\ -2 & \lambda-4 \end{bmatrix}$

$(\lambda-4)^2 - 4 = 0 \Rightarrow \lambda_1 = 2, \ \lambda_2 = 6$

for $\lambda_1 = 2$ $\Sigma v_1 = \lambda_1 v_1$ $\begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 2 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \Rightarrow x_1 + x_2 = 0$

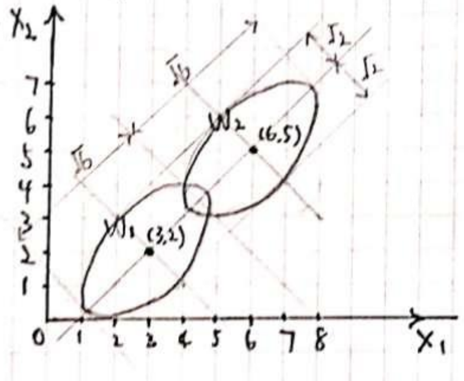
$\therefore v_1 = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{bmatrix}$

for $\lambda_2 = 6$ $\Sigma v_2 = \lambda_2 v_2$ $\begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 6 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \Rightarrow x_1 - x_2 = 0$

$\therefore v_2 = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}$

thus, the orientations of the axes correspond to the two eigenvectors

$$\frac{\text{Length}_{major}}{\text{length}_{minor}} \propto \frac{\sqrt{\lambda_2}}{\sqrt{\lambda_1}}$$



Q2 the decision rule is given by

$x \to w_i$ if $P(w_i|x) = \max_{j=1}^{2} P(w_j|x) = \max_{j=1}^{2} \frac{P(x|w_j)P(w_j)}{P(x)}$

$\therefore$ at decision boundary $P(w_1|x) = P(w_2|x)$

$P(x|w_1)P(w_1) = P(x|w_2)P(w_2)$

substitute the $P(x|w_i)$ with Gaussian distribution and take ln of both sides

$-\frac{1}{2}[\ln|\Sigma_1| + (x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1)] + \ln(P(w_1)) = -\frac{1}{2}[\ln|\Sigma_2| + (x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2)] + \ln(P(w_2))$

since $\Sigma_1 = \Sigma_2 = \Sigma$, $P(w_1) = P(w_2) = \frac{1}{2}$

$(x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1) = (x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2)$

$x^T\Sigma^{-1}x - 2\mu_1^T\Sigma^{-1}x + \mu_1^T\Sigma^{-1}\mu_1 = x^T\Sigma^{-1}x - 2\mu_2\Sigma^{-1}x + \mu_2^T\Sigma^{-1}\mu_2$

$(\mu_1-\mu_2)^T\Sigma^{-1}x + \frac{1}{2}(\mu_2^T\Sigma^{-1}\mu_2 - \mu_1^T\Sigma^{-1}\mu_1) = 0$

$\therefore$ the decision boundary is $w^Tx + c = 0$, where $w = \Sigma^{-1}(\mu_1-\mu_2)$, $c = \frac{1}{2}(\mu_2^T\Sigma^{-1}\mu_2 - \mu_1^T\Sigma^{-1}\mu_1)$

substitute the μ_1, μ_2 and Σ with their values,

$w = \begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix}^{-1} \begin{bmatrix} 3-6 \\ 2-5 \end{bmatrix} = \frac{1}{16-4}\begin{bmatrix} 4 & -2 \\ -2 & 4 \end{bmatrix}\begin{bmatrix} -3 \\ -3 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} \\ -\frac{1}{2} \end{bmatrix}$

$\bar{w} = \frac{w}{|w|} = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}$

$c = \frac{1}{2}\left([6 \;.\; 5]\begin{bmatrix} \frac{1}{3} & -\frac{1}{6} \\ -\frac{1}{6} & \frac{1}{3} \end{bmatrix}\begin{bmatrix} 6 \\ 5 \end{bmatrix} - [3 \;.\; 2]\begin{bmatrix} \frac{1}{3} & -\frac{1}{6} \\ -\frac{1}{6} & \frac{1}{3} \end{bmatrix}\begin{bmatrix} 3 \\ 2 \end{bmatrix} \right) = 4$

$\therefore$ the decision boundary is $x_1 + x_2 - 8 = 0$

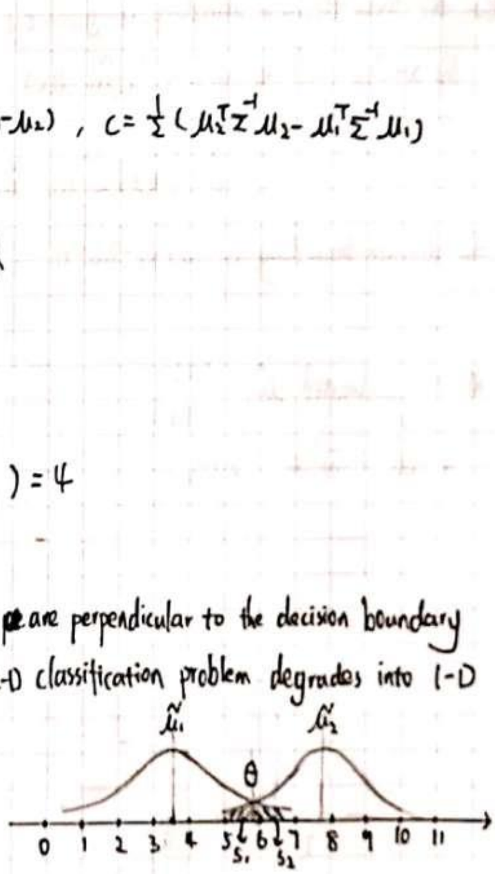
Furthermore, we can find the two long axes coincide and are perpendicular to the decision boundary

if we project these two Gaussian along the vector $\bar{w}$, the 2-D classification problem degrades into 1-D.

$\therefore$ $\tilde{\mu}_1 = \bar{w}^T\mu_1 = [\frac{\sqrt{2}}{2} \; \frac{\sqrt{2}}{2}]\begin{bmatrix} 3 \\ 2 \end{bmatrix} = \frac{5}{2}\sqrt{2}$

$\tilde{\mu}_2 = \bar{w}^T\mu_2 = [\frac{\sqrt{2}}{2} \; \frac{\sqrt{2}}{2}]\begin{bmatrix} 6 \\ 5 \end{bmatrix} = \frac{11}{2}\sqrt{2}$

$\sigma^2 = \bar{w}^T\Sigma w = [\frac{\sqrt{2}}{2} \; \frac{\sqrt{2}}{2}]\begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix}\begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} = 6$



Thus, from the plot of these 1-D gaussians, we can calculate the area of tail of w_1 as

$\theta = \frac{1}{2}(\tilde{\mu}_1 + \tilde{\mu}_2) = 4\sqrt{2}$

$z = \frac{|\theta - \tilde{\mu}_1|}{\sigma} = \frac{|4\sqrt{2} - \frac{5}{2}\sqrt{2}|}{\sqrt{6}} = \frac{\sqrt{3}}{2} = 0.87$

according to the table, $Q(0.87) \approx 0.5 - 0.3078 = 0.1922$

$\therefore$ given x came from class w_1, the probability of error is $\int_{S_2} P(x|w_1)dx = Q(0.87) \approx 0.1922$

due to the symmetry, given x came from class w_2, the probability of error is also 0.1922

the Bayes error is $P(w_1)\int_{S_2} P(x|w_1)dx + P(w_2)\int_{S_1} P(x|w_2)dx = \frac{1}{2} \times 0.1922 + \frac{1}{2} \times 0.1922 = 0.1922$

Q3 (a) the marginal distribution of x_1 and x_2 are

$X_{11} \sim \mathcal{N}(\mu_{11}, \sigma_{11}^2)$ where $\mu_{11}=3$, $\sigma_{11}^2=4$

$X_{12} \sim \mathcal{N}(\mu_{12}, \sigma_{12}^2)$ where $\mu_{12}=2$, $\sigma_{12}^2=4$

the estimated mean

$\hat{\mu}_{11} = E(X_{11}) = \frac{1}{N}\sum_i X_{11}^{(i)} = \bar{X}_{11}$

$\hat{\mu}_{12} = E(X_{12}) = \frac{1}{N}\sum_i X_{12}^{(i)} = \bar{X}_{12}$

the sample variance

$Var(\hat{\mu}_{11}) = \frac{1}{N}\sigma_{11}^2 = \frac{4}{25}$ $\quad Std(\hat{\mu}_{11}) = \sqrt{Var(\bar{X}_{11})} = \frac{2}{5}$

$Var(\hat{\mu}_{12}) = \frac{1}{N}\sigma_{12}^2 = \frac{4}{25}$ $\quad Std(\hat{\mu}_{12}) = \sqrt{Var(\bar{X}_{12})} = \frac{2}{5}$

(b) $\hat{\mu}_{11} = \mu_{11} - std(\hat{\mu}_{11}) = 3 - \frac{2}{5} = 2.6$

$\hat{\mu}_{12} = \mu_{12} - std(\hat{\mu}_{12}) = 2 - \frac{2}{5} = 1.6$

$\therefore \hat{\mu}_1 = \begin{bmatrix} \hat{\mu}_{11} \\ \hat{\mu}_{12} \end{bmatrix} = \begin{bmatrix} 2.6 \\ 1.6 \end{bmatrix}$

thus, the distribution of class w_1 is given by

$\hat{P}(X|W_1) = \mathcal{N}(\hat{\mu}_1, \Sigma)$

the new decision rule is based on this ~~new~~ estimated $\hat{P}(X|W_1)$

i.e. assign $X \to W_1$ if $\hat{P}(X|W_1) > P(X|W_2)$, otherwise assign $X \to W_2$

since the deviation of estimated mean $\hat{\mu}_1$ is along the vector $\bar{w}$, ~~thus~~ the new decision boundary is still a line.

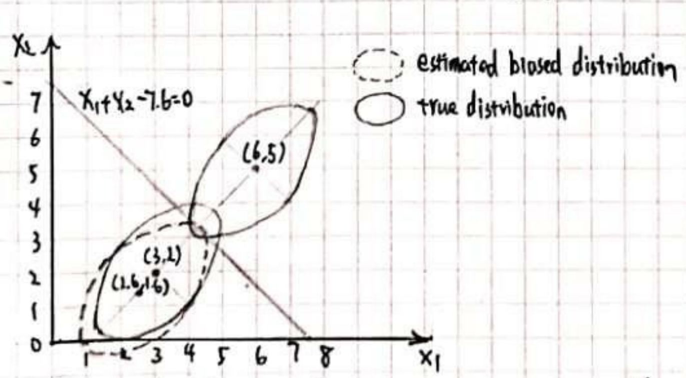
thus, we can say that the x is assigned to the class ~~whose center~~ with closest center mean

(c) the new decision boundary is given by $\hat{w}^T x + \hat{c} = 0$

where $\hat{w} = \Sigma^{-1}(\hat{\mu}_1 - \mu_2)$, $\hat{c} = \frac{1}{2}(\mu_2^T \Sigma^{-1} \mu_2 - \hat{\mu}_1 \Sigma^{-1} \hat{\mu}_1)$

since the boundary is still linear and will pass the central point of $\overline{\hat{\mu}_1 \mu_2}$, which is $(\frac{2.6+6}{2}, \frac{1.6+5}{2})$

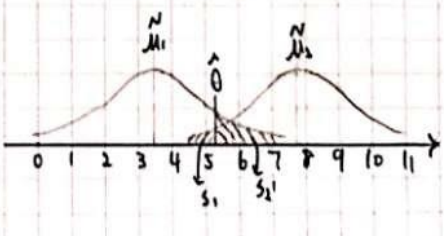
the boundary can be directly ~~given~~ calculated as $x_1 + x_2 - 7.6 = 0$



(d) project the estimated mean along the vector $\bar{w}$

$\tilde{\hat{\mu}}_1 = [\frac{\sqrt{2}}{2} \; \frac{\sqrt{2}}{2}] \begin{bmatrix} 2.6 \\ 1.6 \end{bmatrix} = 2.1\sqrt{2}$

thus, the decision boundary $\hat{\theta} = \frac{1}{2}(\tilde{\hat{\mu}}_1 + \tilde{\mu}_2) = 3.8\sqrt{2}$



Q3(d) cont. given x came from class w_1, the area of tail can be calculated as

$$z_1 = \frac{|\hat{\theta} - \tilde{\mu}_1|}{\sigma} = \frac{|3.8\sqrt{2} - 2.5\sqrt{2}|}{\sqrt{6}} = 0.75$$

according to the table, $Q(0.75) \approx 0.5 - 0.2734 = 0.2266$

thus, given x came from class w_1, the probability of error is $\int_{S_2'} P(x|w_1)dx = 0.2266$

given x came from class w_2, the area of tail can be calculated as

$$z_2 = \frac{|\hat{\theta} - \tilde{\mu}_2|}{\sigma} = \frac{|3.8\sqrt{2} - 5.5\sqrt{2}|}{\sqrt{6}} = 0.98$$

according to the table, $Q(0.98) \approx 0.5 - 0.3365 = 0.1635$

thus, given x come from class w_2, the probability of error is $\int_{S_1'} P(x|w_2)dx = 0.1635$

the ~~[illegible]~~ expected error is $P(w_1) \cdot \int_{S_2'} P(x|w_1)dx + P(w_2) \cdot \int_{S_1'} P(x|w_2)dx = \frac{1}{2}(0.2266 + 0.1635) = 0.1951$

Q4. error indicator is given by $\eta(x) = \begin{cases} 0 & \text{if } x \text{ is assigned to the correct class} \\ 1 & \text{if } x \text{ is assigned to the incorrect class} \end{cases}$

$\therefore$ the expected value of error indicator is

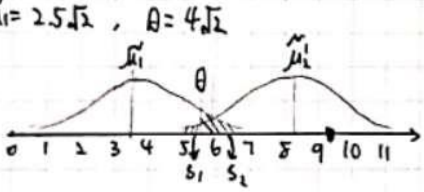
$$E(\eta(x)) = P(w_1)\left(\int_{S_1} P(x|w_1)dx \cdot 0 + \int_{S_2} P(x|w_1)dx \cdot 1\right) +$$
$$P(w_2)\left(\int_{S_2} P(x|w_2)dx \cdot 0 + \int_{S_1} P(x|w_2)dx \cdot 1\right)$$
$$= P(w_1)\int_{S_2} P(x|w_1)dx + P(w_2)\int_{S_1} P(x|w_2)dx$$

(a) $\mu_2^1 = \begin{bmatrix} 6+\frac{2}{5} \\ 5+\frac{2}{5} \end{bmatrix} = \begin{bmatrix} 6.4 \\ 5.4 \end{bmatrix}$

$\tilde{\mu}_2^1 = \left[\frac{\sqrt{2}}{2} \ \frac{\sqrt{2}}{2}\right]\begin{bmatrix} 6.4 \\ 5.4 \end{bmatrix} = 5.9\sqrt{2}$

~~[illegible]~~

$\tilde{\mu}_1 = 2.5\sqrt{2}$, $\theta = 4\sqrt{2}$



$z_1 = \frac{|\theta - \tilde{\mu}_1|}{\sigma} = \frac{|4\sqrt{2} - 2.5\sqrt{2}|}{\sqrt{6}} = 0.87$ $\quad Q(0.87) = 0.5 - 0.3078 = 0.1922$

$z_2 = \frac{|\theta - \tilde{\mu}_2^1|}{\sigma} = \frac{|4\sqrt{2} - 5.9\sqrt{2}|}{\sqrt{6}} = 1.10$ $\quad Q(1.10) = 0.5 - 0.3413 = 0.1587$

given test example X came from class w_1, the probability of error is $\int_{S_2} P(x|w_1)dx = Q(0.87) = 0.1922$

given test example X came from class w_2, the probability of error is $\int_{S_1} P(x|w_2)dx = Q(1.10) = 0.1587$

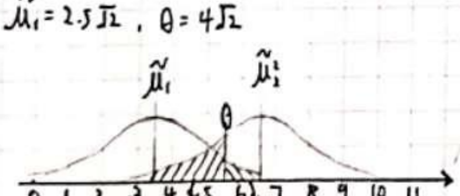
$\therefore$ the expected value of error indicator

$E(\eta(x)) = P(w_1)\int_{S_2} P(x|w_1)dx + P(w_2)\int_{S_1} P(x|w_2)dx = \frac{1}{2} \times 0.1922 + \frac{1}{2} \times 0.1587 = 0.1755$

(b) $\mu_2^2 = \begin{bmatrix} 6-2\times\frac{2}{5} \\ 5-2\times\frac{2}{5} \end{bmatrix} = \begin{bmatrix} 5.2 \\ 4.2 \end{bmatrix}$

$\tilde{\mu}_2^2 = \left[\frac{\sqrt{2}}{2} \ \frac{\sqrt{2}}{2}\right]\begin{bmatrix} 5.2 \\ 4.2 \end{bmatrix} = 4.7\sqrt{2}$

$\tilde{\mu}_1 = 2.5\sqrt{2}$, $\theta = 4\sqrt{2}$



$z_1 = \frac{|\theta - \tilde{\mu}_1|}{\sigma} = \frac{|4\sqrt{2} - 2.5\sqrt{2}|}{\sqrt{6}} = 0.87$ $\quad Q(0.87) = 0.5 - 0.3078 = 0.1922$

$z_2 = \frac{|\theta - \tilde{\mu}_2^2|}{\sigma} = \frac{|4\sqrt{2} - 4.7\sqrt{2}|}{\sqrt{6}} = 0.40$ $\quad Q(0.40) = 0.5 - 0.1544 = 0.3456$

given test example x came from class w_1, the probability of error is $\int_{S_2} P(x|w_1)dx = Q(0.87) = 0.1922$

given test example x came from class w_2, the probability of error is $\int_{S_1} P(x|w_2)dx = Q(0.40) = 0.3456$

$\therefore$ the expected value of error indicator

$E(\eta(x)) = P(w_1)\int_{S_2} P(x|w_1)dx + P(w_2)\int_{S_1} P(x|w_2)dx = \frac{1}{2} \times 0.1922 + \frac{1}{2} \times 0.3456 = 0.2689$

(c) suppose the test set is unbiased as the size of the test is large enough. According to the formula $Var(\mu) = \frac{1}{N}\sigma^2$, if N increases, the variance of the mean of the test set decreases. Thus, the sample mean of the test set will converge to the actual mean if there are enough number of test samples. The closer the sample mean to the actual mean, the better the expected error in test set reveals the performance of the classifier.

Q5. Given the two class distribution, Bayes Error is the optimal (minimal) system error the classifier can get. It is achieved by applying Bayesian decision rule to maximize the probability of $P(w_i|x)$, given the two actual data distribution. Using a biased estimated mean leads to the moving of the decision boundary. If the boundary moves towards one class, the error in that class will increase, while the error in the other class will decrease. And vice versa. But the sum of the error in two classes will be higher than the Bayes error.

If the test set is unbiased, the error in test set should be consistent with the error in the training samples. Suppose one mean in the test set is known and unbiased, if we move the other mean towards the decision boundary, the error in that class will increase. Since the error in the unbiased class is unchanged. The total expected error will increase. If we move the mean against the decision boundary, the error in that class will decrease. Thus the overall expected error will decrease. Therefore, the overall expected error can be either higher or lower than the ideal one, if we use a biased test set. In these cases, the error in the test set cannot properly indicates the performance of the classifier.